

Automatic Assessment of Performance of Hospitals using Subjective Opinions for Sentiment Classification

Muhammad Badruddin Khan

Information Systems Department, College of Computer and Information Sciences
Al Imam Mohammad ibn Saud Islamic University (IMSIU)
Riyadh, KSA

Abstract—Social media is the venue where the opinions are shared in form of text, images and videos by public. Hospitals' performance can be judged by opinions that are written by patients or their relatives. Machine learning techniques can be used to detect sentiments of the opinion givers. For the research work presented in this article, opinions for few big hospitals were collected using Facebook, twitter and hospitals' webpage. The corpus was constructed and the sentiment analysis was performed after few preprocessing tasks. Resources like Stanford POS Tagger and WordNet were used to discover aspects. In this paper, the challenges of annotation of subjective opinions are discussed in detail. Two sentiment lexicons namely NRC-Affect-Intensity lexicon and SentiWordNet 3.0 lexicon were used to calculate sentiment scores of the comments that were used by different machine learning classifiers. Moreover, the results of the experiments on the constructed dataset are provided. For the experiments that aimed to discover overall sentiments of user towards hospital, Random forest outperformed other classifiers achieving accuracy of 76.49% using scores from NRC-Affect-Intensity lexicon. For the experiments that were directed towards discovering sentiments of users towards particular aspect of a hospital, Random forest overtook other classifiers reaching accuracy of 80.7339 % using NRC-Affect-Intensity lexicon sentiment scores. The research results show that machine learning can be very helpful in identifying sentiments of users from their textual comments that are vastly available on different social media platforms. The results can be helpful in improvement of hospital performance and are expected to contribute to growing field of health informatics.

Keywords—Health informatics; Classification Algorithms; Sentiment Analysis; Sentiment Lexicons; Text Mining

I. INTRODUCTION

There are numerous social networking websites such as, Facebook, Google Plus, Twitter, LinkedIn and etc. that have used information technology to contract this globe into a village. People connect to each other and share their opinions, emotions and sentiments in the form of posts and comments using various social networking websites. These posts and comments are valuable source of data that is growing at unprecedented rate. This huge data contains lot of hidden insights, which needs application of data/text mining techniques to be revealed. Education and health are the two most important sectors for the society. A person's deterioration of health affects entire family. Hospitals are the places where

patients come with expectations to restore their health. The services provided by hospitals become part of their experiences. Social media is one of the means to make these good/bad experiences visible to the world. The experiences are shared in different forms. One can write blog post(s), share picture(s)/video(s) and compose comment(s) and these shared opinions act as a trigger and attract more people to share their own personal experiences. These personal experiences can be helpful and beneficial to hospitals' administration and based on opinions of their patients, they can take steps to improve different aspects of their hospitals. Moreover, those patients who plan to receive services of particular hospital in future, can see reviews from previous patients of that particular hospital and decide whether to go to such hospital or not. Machine learning algorithms can be helpful for the task of automatic analysis of such opinions and reviews.

In this paper, personal experiences shared in the form of posts and comments, were used to determine sentiments of people who received medical services from the hospital. Text mining techniques and different sentiment lexicons were used to discover sentiments of experience sharers and opinion givers. The positive side of involvement of machine learning technique to accomplish the task is that machines are expected to be unbiased and unbiased discovery of sentiments can be a useful asset for hospitals to understand their current situations and improve their future performances.

The text of opinions or comments that are shared on social media is not simple. Sometimes, it is even difficult for humans to understand the correct meaning of the comment. Moreover, the granularity level of sentiment in a comment also varies. It means that the text of comment does not always talk about the overall performance of the hospitals with sentences like: "This hospital is good" and "That hospital is not good". The opinion-sharer or commenter can share his/her sentiment about particular aspect of a hospital. It is possible that a hospital performs well with respect to one performance criterion, but with respect to another criterion/criteria, people are not happy with it. These criteria and sentiments related to them are needed to be identified automatically. Aspect based sentiment analysis seeks to understand sentiments about different aspects for specific entities. In this work, the entities are the hospitals and aspects are their different performance criteria.

In this work, comments regarding performances of few big non-Government hospitals of Pakistan were collected. The reason to select non-Government hospitals was that in such hospitals, the patients and/or their relatives pay directly to hospitals (from their own pocket in most of the cases) for health care services and hence their expectation level is high. The patients and/or relatives evaluate the health care services provided by a hospital in terms of the amount that is paid by them. Usually such patients and their relatives are educated and can raise their voice in social media world.

Online comments were gathered from different social media platforms. The step of gathering of comments was followed by laborious manual task of reading each comment and assignment of class to it. Text mining techniques were applied on the built corpora and the results were analyzed.

The rest of paper is organized as follows:

After the introductory section, literature review is presented in order to introduce reader to academic activities similar and related to the work under discussion. Section III titled "Data Preparation" carries full description of the challenges of construction of corpora that can be used as an input to discover sentiments automatically. Section IV discusses results and intuitive reasoning behind the gained results. The evaluation performance of text mining techniques is also given. Section V concludes this paper along with discussion of future research directions.

II. LITERATURE REVIEW

This section will discuss few of the attempts of research that are made in the field of sentiment analysis till now. After describing these attempts, application of machine learning techniques in the field of health care will be discussed. In this regard, similar works will be also mentioned.

The huge amount of text data available at social media sites provides a great opportunity to individuals as well as different groups. The text data is a mixture of facts and opinions. Even though fake facts and forged opinions exist in the cyber world and it is very difficult to quantify extent of fakeness of internet, the worth of available remaining genuine data cannot be denied. The scope of the paper does not allow the author to discuss this issue further, but several researchers like [1], [2] have made efforts in this area of research. Even after subtracting fake content, things do not become simple. Another issue is fact-opinion mixed content. It is difficult for humans to separate facts from opinions from the text content where opinions are mixed with facts. A biased news available on the print media or social media is an example of fact-opinion mixed content. This issue also does not come in the scope of this paper.

The social media platforms provide opportunity to their users to share their opinions and provide their comments on different issues. The area of webpages where these comments or opinions are written and made available to public, become source of almost-pure opinions (it should be noted that facts are sometimes described in opinion sections) that are precious resources for academia as well as public and private sector. Politicians can find public opinion about different political issues from it. Industry can discover their customers' review in

it. Academia can use this data for research purpose. Since the available data is huge, and to deploy human resource to read and summarize these opinions is expensive therefore demand for sentiment-aware applications is great. Nobody from the field expects the machines to be 100% accurate, but even if they are able to produce near-accurate results, it will be enough for decision makers to understand and judge the situation in light of public mood.

Numbers of researchers have conducted researches in order to find the popularity of subject and sentiment analysis, as it is really useful for masses, companies and corporates. Through sentiment analysis, companies can plan for improving themselves and masses can have more insights. If the sentiment analysis is performed taking care of popularity of the subject, then it will be more useful.

In the research world, the notion "sentiment analysis" was firstly used by [3] and other similar term "opinion mining" was first coined by [4]. But the research on the same topic was already started few years ago by [5]–[7].

Document level sentiment analysis was performed by number of researchers including [8], [9]. In the document level sentiment analysis, it is assumed that whole document expresses opinion for single entity. In order to find sentiments at finer level, there was a boom in the field of sentence level sentiment analysis and the main objective of those researches was to find out the sentiment of the each sentence of whole document, which was performed by number of researchers like [10]–[12]. Some work has been done in the field of comparative opinion mining and [13] has done research on YouTube comments for the same purpose. Even in a sentence with single entity, there can be aspect(s) with respect to which there exist positive sentiment(s) and with respect to other aspect(s), the negative sentiment(s) can also exist. Sentence level sentiment analysis provide overall sentiments at sentence level but not at aspect level. Therefore, aspect level sentiment analysis was introduced, which was earlier called as feature level sentiment analysis. The researchers in [14] discussed the usage of state-of-art techniques of CNN and LSTM for the purpose of aspect-level opinion mining. The authors in [15] presented the same issue in context of recommender systems comparison. Many researches on aspect-oriented sentiment analysis were performed via different methods. The investigators in [16] found aspects by frequency of nouns in the whole document and then performed sentiment analysis on retrieved aspects. Author in [17] found relatedness of noun and adjective and via this method tried to retrieve aspect.

III. DATA PREPARATION

Social media platform was the source of data that was input for this research work. The experiences of patients and their relatives regarding health care services, if shared on social media, can be found online at no cost. Through these types of comments, other people can have an idea of hospital's performance. Hospital administration can also use these comments to improve their services, thus enabling them to achieve satisfaction of their patients and attendants. However, there is no straight forward way of achieving this goal because some people post irrelevant comments. For example, they start marketing or branding their products, or they start to post

jokes. Such type of comments becomes noise that need to be tackled during preprocessing step. On the other hand, there are relevant comments, which are related to topic. However, the relevant comments come in different varieties. There are various types of relevant comments that were discovered during process of formulation of dataset. Following categories can be constructed after careful study of users' relevant comments:

- A. Direct comments with opinion (pointing to topic with opinion)
- B. Direct comments without opinion (pointing to topic without an opinion)
- C. Informative comments (provide more information on topic)
- D. Informative comments with aspects (provide information about aspects of topic)
- E. Comparative comments (topic-level comparison)
- F. Comparative comments with aspects (aspect-level comparison)
- G. Declarative comments.

In this research, the main objective is to **discover** aspect(s) from the comments and then based on extracted aspects, assignment of sentiment scores (positivity, negativity, or objectivity) is needed to be performed. It is possible that in one comment, commentator gives his/her sentiments with respect to multiple aspects. For such type of comments, the need is to extract all the aspects from the comments and then to assign sentiment score based on every extracted aspect. Following two comments are given as an example. The hospital names are replaced by X and Y. It should be noted that the two comments presented as example in following lines, are public comments that are exactly copied here and hence spelling and grammatical mistakes can be found. However, hospital names are replaced. The two example comments are as follows:

1. X hospital environment is pretty good, whereas administration are irresponsible.
2. Y nurses are comparatively helpful than X nurses but its parking is very conusted, particularly for bikers.

In the first comment, topic or entity is X hospital and aspects are environment and administration. Comment is positive with respect to environment, whereas it is negative in the case of administration. Second comment is not only aspect-oriented but also comparative in nature. The commenter of second comment began with comparison of the nursing service of X hospital and Y hospital and with respect to nursing, comment is positive for Y hospital and negative for X hospital. In the second part of the second comment, entity is "Y hospital" and the aspect in discussion is parking and with respect to "parking" aspect, comment is negative for "Y hospital".

The second example comment is an example of comparative comment with aspects. This research work addresses comments that have/have not aspect(s) based

sentiments but inter-hospital comparison does not exist in them.

At the beginning of research, around 10,000 reviews were fetched but all of these reviews/comments were not in textual form. For example, most of the people gave their opinion on the basis of **stars** in comments. For example, 5 stars means "I love this" and 1-star means "I hate this". Such type of reviews was irrelevant for the research. After carefully reading and pruning the comments, less than thousand comments were left that became the subject of the study.

In order to construct dataset consisting of aspect-oriented comments, following tasks were performed:

- A. Comments fetching
- B. Wiping out noisy and irrelevant comments

In the upcoming sub-sections, discussion about the above tasks will be presented followed by description of process of removing inconsistencies from comments, and after that problem with annotations of comments will be discussed in detail. Experiments, discussions, conclusion and future research will be discussed afterwards.

Before you begin to format your paper, first write and save the content as a separate text file. Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

A. Comments Fetching

The first and the foremost step was to gather the comments or reviews of the people. Graph API provided by Facebook, was used to fetch Facebook comments as well as reviews. On the other hand, Twitter API was also used to get tweets of people. However, only few tweets contained discussion of the performance of hospitals and most of the data was fetched via Facebook which was nearly 10000 reviews/comments that was reduced later to less than one thousand comments due to the research work domain constraint. Both providers gave feeds/tweets in various forms. The data was fetched in JSON format and after some processing, was stored in the csv file. To perform these operations, a program was written in JAVA, using which comments were fetched in JSON format and the fetched data was provided to GSON converter (Library written by Google. Inc.), which converted JSON into plain java object (POJO). POJO was read line by line by the program and the data from it was inserted into the CSV file. For the multi-line comments, end of line character was replaced by the space so that every comment fit into single line. The data from hospital's official review page was also fetched.

B. Wiping Out Noisy and Irrelevant Comments

Data which was collected in the first step was not in the usable form and there was irrelevant data also. It is a common practice in Facebook that people tag their friends by typing

their names in the comment therefore dataset contained lot of comments and reviews in which names of people were present. Such data was removed manually after reading all comments. Regarding data that was fetched from official hospital's review page, another difficulty and limitation was faced. In those pages, it was not mandatory for people to provide reviews in form of narrations as the field on the form of the webpage was optional for them to fill. The mandatory thing was that they have to give star(s) to provide their feedback, which was not useful for the purpose of this research. Hence such star reviews were also discarded from the input data file. Moreover, there were some comments, which were not relevant. Some people gave marketing comments and some posted jokes. Such comments were also removed manually after reading them. Furthermore, there were some comments, which were written in Roman Urdu language (Urdu language written with the Roman script) to represent opinions in Urdu language. In order to avoid complexity, such comments were also truncated from the CSV file. After above-mentioned preprocessing steps, dataset was ready to be used to perform aspect-based sentiment analysis.

C. Class and Aspect Assignment

For supervised learning, labelled dataset is required that can be used by different classifiers to construct the model that can automatically perform sentiment analysis. In order to label records of the dataset, manual annotation was performed for assignment of class and aspects to the comment/review.

Class assignment was relatively easier than aspect assignment because the whole comment or review was only to be assigned the label of positivity, negativity or objectivity. Whereas, for aspect assignment whole comment was needed to be read to discover aspect(s) after understanding the context and then step of assignment of class (i.e. sentiment) based on discovered aspect, was performed. There were some comments with more than one aspect. In the constructed dataset, maximum of three aspects in a single comment exist. In next lines, few examples are provided to show the complexity of the problem. It should be noted that spelling and grammatical errors can be found in the provided examples. Hospital names in the comments are replaced by symbols X and Y.

1) Class Assignment for Whole Comment

Three classes are usually assigned in normal sentiment classification. From the constructed comments list, example of each class is given below:

a) *Positive (Pos)*: Keep up the good work and pls make better parking arrangements

b) *Negative (Neg)*: Tests is expensive of X Hospital. Normally are twice expensive than any other lab in the city

c) *Neutral (Neu)*: We have to trust doctors even some are bad and some good too.

D. Aspect-based Class Assignment

As discussed earlier, one can provide opinion with respect to more than one aspect in a single comment or review. The aspects are needed to be discovered in the first step followed by the step of sentiment assignment to discovered aspects.

Annotation process of assignment of sentiments to aspects recorded class values using following taxonomy:

<Aspect1>_<class>-<Aspect2>_<class>-
<Aspect3>_<class>

Some examples of aspect-based class assignments are given below:

a. Single Aspect Comment Example

In the following example, the commenter has given opinion about quality of healthcare services. The text of the comment is as follows:

I have never ever seen such type of quality healthcare services. Simply outstanding...

Here aspect is "services" and assigned class is **positive** hence the label based on used taxonomy will be "services_positive".

b. Double Aspect Comment

In the following example, the commenter has given opinion about performance of hospital with respect to different aspects. The text of the comment is as follows:

I consider X hospital to be a hospital full of unprofessional doctors and nurses. You have to micro manage doctors and nurses. Unless you request something (Paging or requesting a doctor) 2-3 times, it won't happen. Dr. XX on Special care unit in private section on 3rd floor is the most phathtic and unprofessional doctor I have ever met. He clearly does not like his job. We plan to sue X hospital of all the neglect they are doing to our father and I'll make sure that I speak the true colors of X in social media in near future to come

Here commenter is complaining about the service of doctor and nurses at X Hospital. For the above comment, the first aspect is doctor and the second aspect is nurse. Hence label will be "doctors_negative-nursing_negative"

Table I presents few aspects that were present in comments of the constructed dataset.

TABLE I. LIST CONTAINING FEW ASPECTS PRESENT IN DATASET

Nursing	Cleaning	Parking
Treatment	Doctors	Care
Environment	Food	Patient
Cafeteria	Facilities	Diagnosis
Management	Expense	Discount

E. Difficulties in Assignments of Sentiments

Annotating sentiments with respect to multiple aspects is marginally difficult than annotating sentiment for entire comment. Various difficulties were faced when annotating sentiments with respect to aspects and entire comment. Some of them are given in following points:

1) It is difficult to understand the polarity of the sentences as well as aspects due to poor usage of English grammar.

2) Too much typos can be present in number of comments.

3) Existence of ambiguity in the comments. For example, the following comment provides the insights to approximate the extent of the problem of ambiguity.

Comment: "Impact of doctor is gt than other and impact of nurse is gt then doctors. It is good for us that sm doctor r gud in Y hospital but we cant do ny thng for bad doc".

The above comment is copied from the comments list. It is really difficult even for human being to understand on which side, the polarity of the commenter is. This is the comment which has ambiguity, spelling mistakes and poor usage of English at the same time.

There are two different type of ambiguities in reviews or subjective opinions.

Ambiguity Type – 1:

There were some comments that contained ambiguous statements and it was hard to decide the sentiments of such comments and as a result difficulties were faced while annotating such comments.

Ambiguity Type – 2:

Some comments were written in way that punctuation and grammatical errors and typo mistakes created the impression of presence of ambiguity in them. Hence such comments were apparently ambiguous.

Above example contains both type of ambiguities as it is really hard to decide polarity and there are too many typo and other mistakes in that comment.

F. Finding Aspects

There are four different methods to find the aspect from the text.

- 1) Extraction based on frequent nouns and noun phrases
- 2) Extraction by exploiting opinion and target relations
- 3) Extraction using supervised learning
- 4) Extraction using topic modeling

The simplest method is method number 1. In this research work, the first method with some modification was used to extract aspects. Custom logic was developed to overcome

different problems associated with finding aspects. The algorithm was able to fetch significant number of aspects like doctors, treatment, cafeteria, staff, parking and quality. The algorithm was unable to find few aspects like facilities, nursing, care and management due to low number of comments with such aspects. Algorithm also made some errors in identifying non-aspects as aspects.

G. Sentence Level Sentiment Classification

WEKA was used to perform machine learning task. 10-folds cross-validation was used to test different machine learning algorithm results on the constructed dataset. Number of classes was three namely positive, negative and neutral. Two sentiment lexicons were used to provide sentiment scores of each comment.

H. Aspect-based Sentiment Classification

Two lexicons were used along with different classifiers for aspect-based sentiment classification. 10-folds cross-validation was used in test settings for different experiments. The experiments aimed to find the sentiment of users towards particular aspect of performance of a hospital. Special program was built to extract neighboring words as tokens that were later merged to form new concise comment. Number of classes was two namely positive and negative.

IV. RESULTS AND DISCUSSION

A. Sentence Level Sentiment Classification

Experiments were performed with different settings in Weka environment using package for analyzing Affect in tweets[18] and the results of the experiments for sentence level sentiment classification using two lexicons are presented in Table II and Table III. It should be noted that the application of lexicons on the dataset in preprocessing stage resulted in generation of new attributes that carried different scores for comments or tweet. These new attributes were used for classification using different classifiers. Moreover, no tokenization was performed and it was tested that how newly generated attributes help in the sentiment classification process. For example, when lexicon NRC-Affect-Intensity lexicon [19] was applied on the dataset, new attributes that were generated were as follows:

TABLE II. RESULTS OF EXPERIMENTS WITH COMBINATION OF NRC-AFFECT-INTENSITY LEXICON AND DIFFERENT CLASSIFIERS

Classifier	Lexicon applied in preprocessing stage	Class	Accuracy	Precision	Recall
Decision Tree J48	NRC-Affect-Intensity lexicon [19] No tokenization	Positive	72.6098 %	0.524	0.418
		Negative		0.784	0.896
		Neutral		0.444	0.200
Naïve Bayes	NRC-Affect-Intensity lexicon [19] No tokenization	Positive	47.2868 %	0.313	0.063
		Negative		0.729	0.541
		Neutral		0.192	0.825
Random Forest	NRC-Affect-Intensity lexicon [19] No tokenization	Positive	76.4858 %	0.587	0.468
		Negative		0.815	0.918
		Neutral		0.591	0.325

TABLE III. RESULTS OF EXPERIMENTS WITH COMBINATION OF SENTIWORDNET LEXICON AND DIFFERENT CLASSIFIERS

Classifier	Lexicon applied in preprocessing stage	Class	Accuracy	Precision	Recall
Decision Tree J48	SentiWordNet 3.0 [20] No tokenization	Positive	71.8346 %	0.531	0.329
		Negative		0.748	0.940
		Neutral		0.000	0.000
Naïve Bayes	SentiWordNet 3.0[20] No tokenization	Positive	56.8475 %	0.100	0.013
		Negative		0.737	0.731
		Neutral		0.207	0.575
Random Forest	SentiWordNet 3.0[20] No tokenization	Positive	68.9922 %	0.500	0.392
		Negative		0.771	0.866
		Neutral		0.167	0.100

NRC-Affect-Intensity-anger_Score, NRC-Affect-Intensity-fear_Score, NRC-Affect-Intensity-sadness_Score, NRC-Affect-Intensity-joy_Score.

Table II clearly demonstrates that Naïve Bayes is not a suitable classifier to be used when there is no tokenization. Ensemble method of Random Forest outperformed decision tree as expected. Table III shows that SentiWordNet lexicon application on the dataset followed by different classifier usage was not promising as compared to NRC-Affect-Intensity lexicon. Even though decision tree outperformed Random Forest however it can be clearly seen that for Neutral comments, decision tree classification model had no clue for detection of neutral comments. Naïve Bayes performance saw some improvement for the SentiWordNet lexicon as compared to NRC-Affect-Intensity lexicon.

B. Aspect-based Sentiment Classification

After discovery of aspect, the three neighbor words before and three neighbor words after aspect term were taken as the input for the experiment. For example, the comment “X is the best hospital and especially X nursing is the excellent”, has the aspect of “nursing” under discussion. After preprocessing, the

processed comment for experiment was “and especially X is the excellent”. The aspect term “nursing” was removed from the comment and three neighbor words before the aspect term and three neighbor words after the aspect term were included for the experiment purpose. For aspect based classification, only positive and negative comments were present in the dataset. Table IV and Table V show the results when the sentiment analysis was applied to discover user sentiments about single aspect. Again no tokenization was performed.

Table IV shows that Random Forest again outperformed other classifiers. Moreover, it can be seen that Naïve Bayes classifier performance has increased as compared to the performance on the full comment. Table V demonstrates an unexpected phenomenon that Naïve Bayes outperformed decision tree and Random Forest classifier. The reason may be the availability of only two scores for the three classifiers and the neighborhood settings for input formulation for experiments of single aspect classification may be more suitable for probabilistic requirements that Naïve Bayes classifier demands. The absence of neutral comments can also be seen as the reason for better performance of Naïve Bayes classifier.

TABLE IV. RESULTS OF EXPERIMENTS FOR ASPECT-BASED SENTIMENT CLASSIFICATION ON COMMENTS DISCUSSING SINGLE ASPECT, WITH COMBINATION OF NRC-AFFECT-INTENSITY LEXICON AND DIFFERENT CLASSIFIERS

Classifier	Lexicon applied in preprocessing stage	Class	Accuracy	Precision	Recall
Decision Tree J48	NRC-Affect-Intensity lexicon [19] No tokenization	Positive	79.8165%	0.727	0.296
		Negative		0.806	0.963
Naïve Bayes	NRC-Affect-Intensity lexicon [19] No tokenization	Positive	75.2294 %	0.500	0.333
		Negative		0.802	0.890
Random Forest	NRC-Affect-Intensity lexicon [19] No tokenization	Positive	80.7339 %	0.750	0.333
		Negative		0.814	0.963

TABLE V. RESULTS OF EXPERIMENTS FOR ASPECT-BASED SENTIMENT CLASSIFICATION ON COMMENTS DISCUSSING SINGLE ASPECT, WITH COMBINATION OF SENTIWORDNET LEXICON AND DIFFERENT CLASSIFIERS

Classifier	Lexicon applied in preprocessing stage	Class	Accuracy	Precision	Recall
Decision Tree J48	SentiWordNet 3.0 [20] No tokenization	Positive	74.3119 %	0.474	0.333
		Negative		0.800	0.878
Naïve Bayes	SentiWordNet 3.0[20] No tokenization	Positive	77.0642 %	0.563	0.333
		Negative		0.806	0.915
Random Forest	SentiWordNet 3.0[20] No tokenization	Positive	74.3119 %	0.478	0.407
		Negative		0.814	0.854

V. CONCLUSION

Health care services can be evaluated by comments present on social media platforms. Text mining techniques enable automatic discovery of sentiments of opinion givers. This paper described the challenges associated with assessment of performances of hospitals using subjective opinion. It discussed the challenges of formulation and annotation of dataset. It presented how different aspects of health care services can be discovered. It provided results of experiments where sentiment analysis was performed on full comments. Moreover, results were also provided for experiments that aimed to discover sentiment of user for particular aspect of the hospital. In experiments aiming to discover the overall sentiment of the user towards hospital, Random forest and Decision tree classifiers provided good results for the NRC-Affect-Intensity lexicon and SentiWordNet 3.0 lexicons. The experiments that were directed toward finding users' opinion about particular aspect of a hospital, special type of preprocessing was done on input comments and the size of input comments was drastically reduced to maximum of 6 words as a heuristic. The results show that Naïve bayes classifier performance increased drastically reaching to 77.06% using SentiWordNet 3.0 scores. Random forest classifier achieved 80.73% accuracy in the experiments using sentiment scores from NRC-Affect-Intensity lexicon.

In this paper, two sentiment lexicons and three classifiers were used with no tokenization. In future, the work will be enhanced in all directions with inclusion of more lexicons and more classifiers in experiments along with tokenization. In this paper, results of experiment to discover sentiments for single aspect in user comments were presented. In future, the results of experiments that aim to discover sentiment towards multiple aspects of hospital in a single comment will be presented. Depending on the availability of data, one of the prospective area for enhancement of the presented research is the domain of comparative opinion mining where user compares the performance of a hospital with another hospital. Further research in this area is also planned so that machine learning performance in this arena is also explored.

REFERENCES

[1] T. Lavergne, T. Urvoy, and F. Yvon, "Detecting Fake Content with Relative Entropy Scoring," PAN, vol. 8, pp. 27–31, 2008.
[2] D. M. Lazer et al., "The science of fake news," Science, vol. 359, no. 6380, pp. 1094–1096, 2018.
[3] T. Nasukawa and J. Yi, "Sentiment Analysis: Capturing Favorability Using Natural Language Processing," in Proceedings of the 2Nd

International Conference on Knowledge Capture, 2003, pp. 70–77, doi: 10.1145/945645.945658.

[4] K. Dave, S. Lawrence, and D. M. Pennock, "Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews," in Proceedings of the 12th International Conference on World Wide Web, 2003, pp. 519–528, doi: 10.1145/775152.775226.
[5] S. R. Das et al., "Yahoo! for amazon: Sentiment extraction from small talk on the web," in 8th Asia Pacific Finance Association Annual Conference, 2001.
[6] S. Morinaga, K. Yamanishi, K. Tateishi, and T. Fukushima, "Mining Product Reputations on the Web," in Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2002, pp. 341–349, doi: 10.1145/775047.775098.
[7] J. Wiebe, "Learning Subjective Adjectives from Corpora," in Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence, 2000, pp. 735–740.
[8] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs Up?: Sentiment Classification Using Machine Learning Techniques," in Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10, 2002, pp. 79–86, doi: 10.3115/1118693.1118704.
[9] P. D. Turney, "Thumbs Up or Thumbs Down?: Semantic Orientation Applied to Unsupervised Classification of Reviews," in Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, 2002, pp. 417–424, doi: 10.3115/1073083.1073153.
[10] V. Hatzivassiloglou and J. M. Wiebe, "Effects of Adjective Orientation and Gradability on Sentence Subjectivity," in Proceedings of the 18th Conference on Computational Linguistics - Volume 1, 2000, pp. 299–305, doi: 10.3115/990820.990864.
[11] W.-H. Lin, T. Wilson, J. Wiebe, and A. Hauptmann, "Which Side Are You on?: Identifying Perspectives at the Document and Sentence Levels," in Proceedings of the Tenth Conference on Computational Natural Language Learning, 2006, pp. 109–116.
[12] E. Riloff and J. Wiebe, "Learning Extraction Patterns for Subjective Expressions," in Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing, 2003, pp. 105–112, doi: 10.3115/1119355.1119369.
[13] A. U. R. Khan, M. Khan, and M. B. Khan, "Naïve Multi-label Classification of YouTube Comments Using Comparative Opinion Mining," Procedia Comput. Sci., vol. 82, pp. 57–64, 2016, doi: http://dx.doi.org/10.1016/j.procs.2016.04.009.
[14] W. Quan, Z. Chen, J. Gao, and X. T. Hu, "Comparative Study of CNN and LSTM based Attention Neural Networks for Aspect-Level Opinion Mining," in 2018 IEEE International Conference on Big Data (Big Data), 2018, pp. 2141–2150.
[15] M. Hernández-Rubio, I. Cantador, and A. Bellogin, "A comparative analysis of recommender systems based on item aspect opinions extracted from user reviews," User Model. User-Adapt. Interact., pp. 1–61, 2018.
[16] B. Liu, M. Hu, and J. Cheng, "Opinion Observer: Analyzing and Comparing Opinions on the Web," in Proceedings of the 14th International Conference on World Wide Web, 2005, pp. 342–351, doi: 10.1145/1060745.1060797.

- [17] S. Somasundaran, J. Ruppenhofer, and J. Wiebe, "Discourse Level Opinion Relations: An Annotation Study," in Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue, 2008, pp. 129–137.
- [18] F. Bravo-Marquez, E. Frank, B. Pfahringer, and S. M. Mohammad, "AffectiveTweets: a Weka Package for Analyzing Affect in Tweets," *J. Mach. Learn. Res.*, vol. 20, no. 92, pp. 1–6, 2019.
- [19] S. M. Mohammad, "Word Affect Intensities," CoRR, vol. abs/1704.08798, 2017.
- [20] S. Baccianella, A. Esuli, and F. Sebastiani, "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining," in Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17-23 May 2010, Valletta, Malta, 2010.